

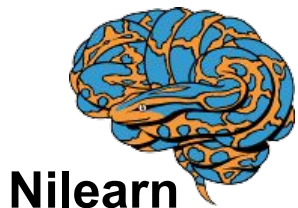
BIDS

BRAIN IMAGING DATA STRUCTURE

Remi Gau

<https://allmylinks.com/remi-gau>

22nd November 2023



BIDS principles

BIDS is:

- Brain Imaging Data Structure
- About [FAIR data](#)
 - Findable
 - Accessible
 - **Interoperable**
 - **Reusable**

Why manage data?

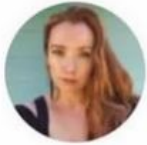
We want:

- answer our scientific question

without spending time :

- organizing data

Why manage data?



Coley

@ColeyTangerina

In every **lab**, there is a person who stacks the **data** like a Scandinavian architect and a person who stacks the **data** like a racoon on meth.

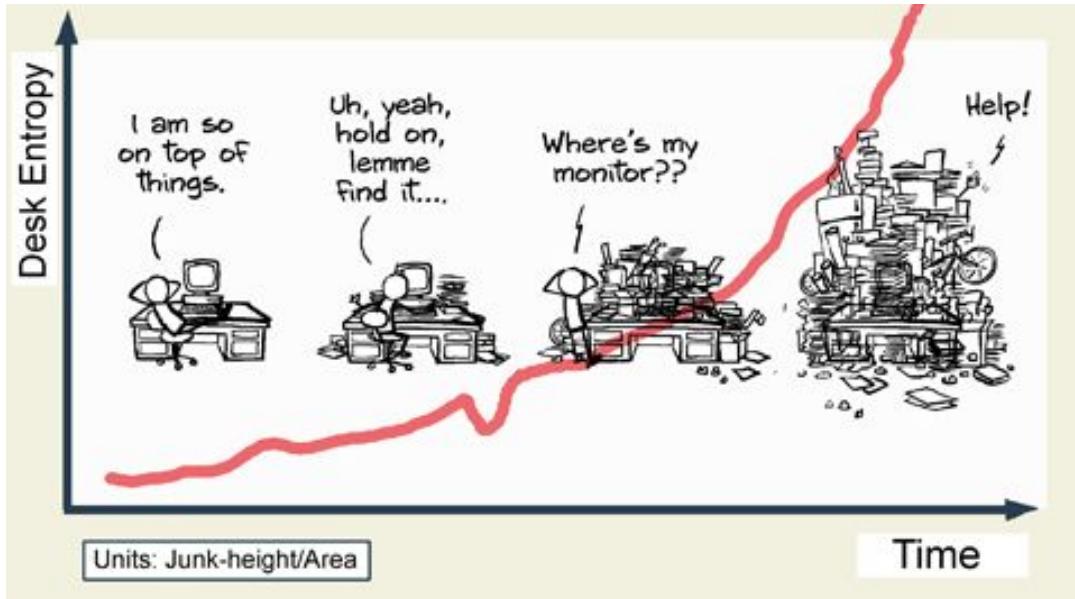
We want:

- answer our scientific question

without spending time :

- organizing data

Why manage data?



Source: [Phd comics](#)

[PhD](#)

We want:

- answer our scientific question without spending time :
- organizing data

Why manage data?

Cleaning your data will make it easier to share them with your closest collaborator

Why manage data?

Cleaning your data will make it easier to share them with your closest collaborator:

- Yourself...
- ... in 6 months.

Why BIDS?

- Data and code managements are not (yet) commonly taught.
- By applying BIDS you are implicitly applying several good data management principles.

Data management principles

Research data management

(explained with pasta)



Source data



Raw data



Derivative data*

* as expected from reading the method section

Research data management

(explained with pasta)

What the dataset actually looks like



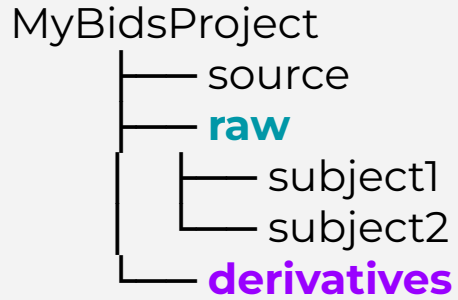
Shared dataset

Research data management (explained with pasta)

What the dataset actually looks like

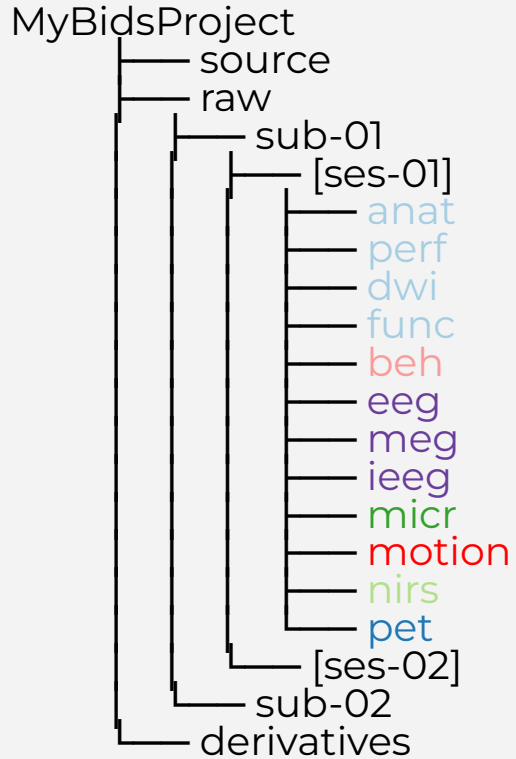


Modularize data

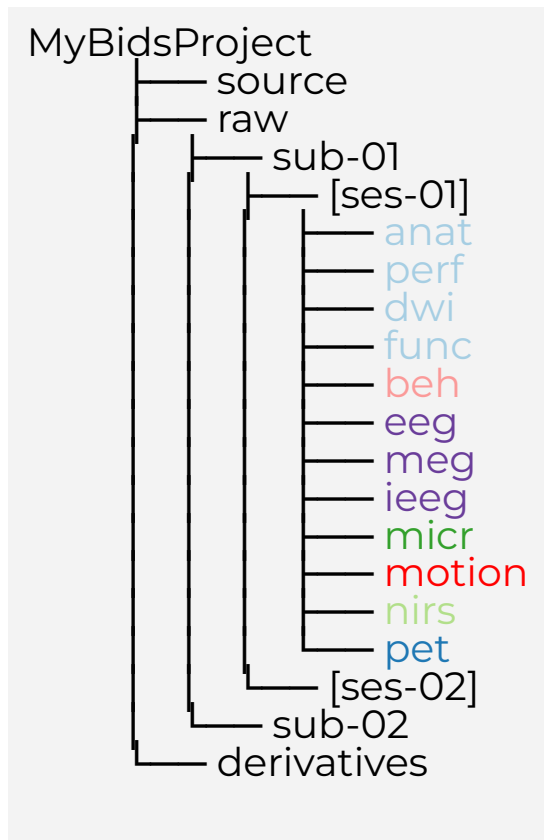


Folder naming

Folder naming



Folder naming



| | modality | | | | | | |
|----------|----------|-----|------|------------|------------|------|--------|
| | MRI | PET | meeg | behavioral | microscopy | NIRS | motion |
| | anat | pet | eeg | beh | micr | nirs | motion |
| datatype | func | | meg | | | | |
| | dwi | | ieeg | | | | |
| | perf | | | | | | |

BIDS extension proposals:

- Animal electrophysiology
- Eyetracking
- Magnetic resonance spectroscopy
- Non invasive brain stimulation
- ...

BIDS file naming

BIDS file naming

key1-value_key2-value_suffix.extension

BIDS file naming

key1-value_key2-value_suffix.extension

- **Suffix** preceded by an **underscore**
- Entities are composed of **key-label** pairs separated by **underscores**
- **Key** and **label** separated by **hyphen**
- **Keys**, **labels**, **suffixes** can only contain letters and / or numbers.

BIDS file naming



Anat filename template

```
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_ce-<label>][_rec-<label>][_run-<index>][_part-<mag|phase|real|imag>]_<suffix>.json
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_ce-<label>][_rec-<label>][_run-<index>][_part-<mag|phase|real|imag>]_<suffix>.nii[.gz]
```

EEG filename template

```
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_run-<index>]_eeg.<extension>
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_run-<index>]_eeg.json
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_run-<index>]_events.json
sub-<label>[_ses-<label>][_task-<label>][_acq-<label>][_run-<index>]_events.tsv
```

File formats

File formats

Use existing open format standards when possible.

File formats

Use existing open format standards when possible.

- MRI
 - Volume: Nifti
 - Surface: Gifti
- PET
 - Nifti
- EEG / MEG / iEEG
 - [European data format](#) (.edf)
 - [BrainVision Core Data Format](#) (.vhdr, .vmrk, .eeg)
 - [EEGLAB](#) (.set)
 - [Neurodata Without Borders](#) (.nwb)
 - ...
- Microscopy
 - ome.tif
 - ome.btf
 - png
 - tif
- Near infra-red spectroscopy
 - snirf

File formats

Prefer formats that can be opened by most softwares, including simple text editors.

File formats

Prefer formats that can be opened by most softwares, including simple text editors.

Stay away from **.doc**, **.xls**, **.mat**, **.pickle...**

File formats: table

- TSV files : Tabulation Separated Values

participants.tsv

```
participant_id\tage\tsex\tspecies\tgroup\thandedness  
sub-01\t31\tM\tHomo sapiens\tcontrol\t0.7  
sub-02\t25\tM\tHomo sapiens\tcontrol\t0.8  
sub-03\t30\tF\tHomo sapiens\tpatient\t0.3
```

File formats: metadata

- JSON files: JavaScript Object Notation
 - for attribute-values pairs
 - as “side-car” to a data file

sub-control01_task-nback_bold.nii
sub-control01_task-nback_bold.json
sub-control01_task-nback_events.tsv
sub-control01_task-nback_events.json

Example:

sub-control01_task-nback_bold.json

```
{  
  "TaskName": "N Back",  
  "RepetitionTime": 0.8  
}
```

Tips

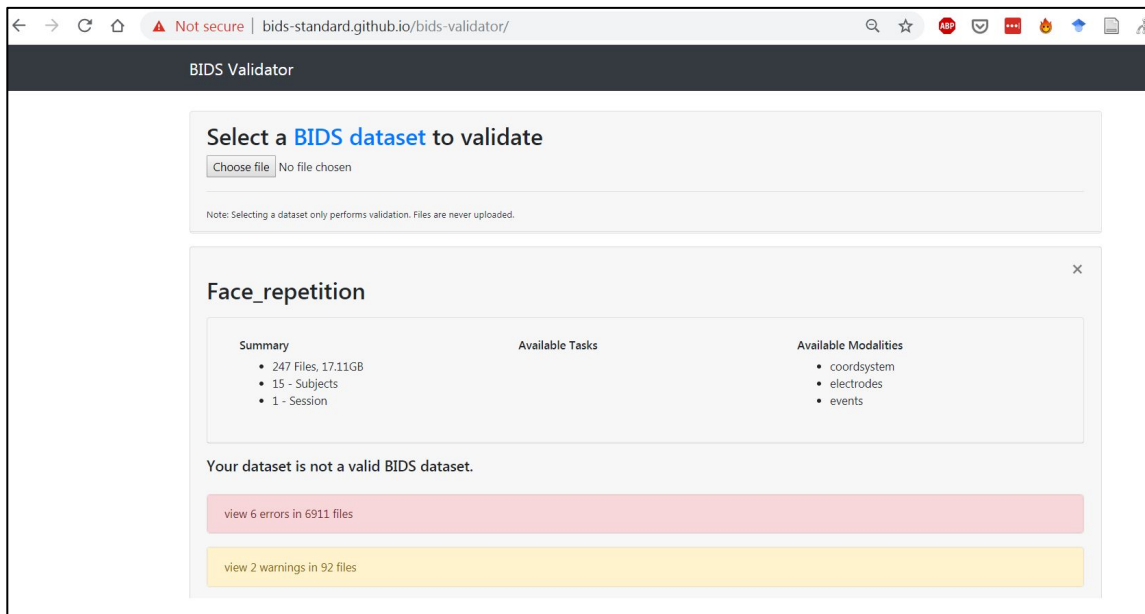
Getting started

- BIDS [starter kit](#)
- youtube channel: [playlists](#)

The screenshot displays the BIDS Starter Kit website. On the left is a navigation sidebar with a search bar and a list of categories: FOLDERS AND FILES, TUTORIALS, and RESOURCES. The main content area is titled 'Welcome to the BIDS Starter Kit' and includes a subtitle 'How to get started with the Brain Imaging Data Structure'. Below this is a paragraph describing the site as a community-curated collection of tutorials, wikis, and templates. A navigation bar contains links for Specification, FAQ, Chat, Forum, Youtube, and Podcast. A video player is embedded, showing a thumbnail for 'What is the BIDS steering group?' with the BIDS logo and the text 'The BIDS steering group' and 'With Kirstie Whitaker & Franklin Feingold'. A 'Watch on YouTube' button is visible at the bottom of the video player. On the right side of the page, there is a 'Contents' menu listing: Motivation, Project Summary, Philosophy, Benefits, Users, and Acknowledgements. The word 'Motivation' is also visible at the bottom of the page.

BIDS validator

- [BIDS validator](#)
 - Web
 - CLI



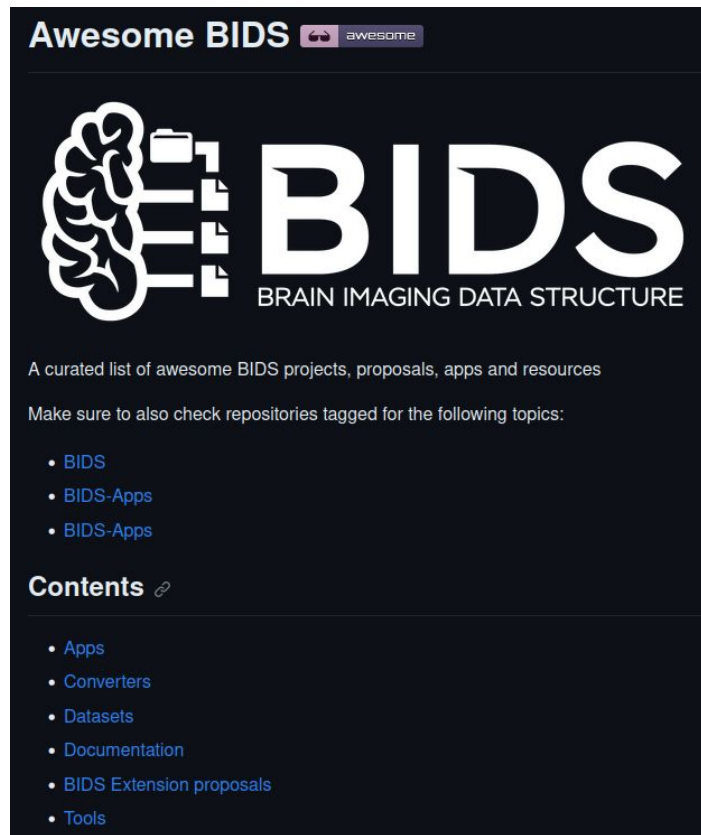
The screenshot shows the BIDS Validator web interface in a browser. The page title is "BIDS Validator". The main heading is "Select a BIDS dataset to validate". Below this is a "Choose file" button and the text "No file chosen". A note states: "Note: Selecting a dataset only performs validation. Files are never uploaded." A modal window titled "Face_repetition" is open, displaying the following information:

| Summary | Available Tasks | Available Modalities |
|--|-----------------|---|
| <ul style="list-style-type: none">• 247 Files, 17.11GB• 15 - Subjects• 1 - Session | | <ul style="list-style-type: none">• coordsystem• electrodes• events |

Below the modal, a message states: "Your dataset is not a valid BIDS dataset." There are two buttons: "view 6 errors in 6911 files" (highlighted in pink) and "view 2 warnings in 92 files" (highlighted in yellow).

List of resources

- [Awesome BIDS](#)



The image shows a screenshot of the 'Awesome BIDS' GitHub repository page. At the top, it says 'Awesome BIDS' with a GitHub logo and a 'awesome' badge. Below this is a large white logo consisting of a brain on the left and the letters 'BIDS' in a large, bold font on the right, with 'BRAIN IMAGING DATA STRUCTURE' written in smaller letters underneath. The text below the logo reads: 'A curated list of awesome BIDS projects, proposals, apps and resources'. It then says 'Make sure to also check repositories tagged for the following topics:' followed by a bulleted list: '• BIDS', '• BIDS-Apps', and '• BIDS-Apps'. Below this is a section titled 'Contents' with a link icon, followed by another bulleted list: '• Apps', '• Converters', '• Datasets', '• Documentation', '• BIDS Extension proposals', and '• Tools'.

BIDS datasets

- [BIDS examples](#)

The screenshot shows the GitHub repository page for 'bids-examples'. The page has a dark blue header with the repository name and a search bar. On the left, there is a navigation menu with 'Introduction' and 'Contributing' links. The main content area features a 'validate_datasets' badge with a 'passing' status, followed by a link to the repository on the BIDS standard website. Below this, the title 'bids-examples' is displayed in a large blue font. The text explains that the repository contains BIDS-compatible datasets with empty raw data files and lists two purposes: writing software tests and serving as a structural example. A bold warning states that all raw data files are empty. Finally, it notes that some metadata headers are still intact in certain files.

bids-examples

The BIDS examples

Introduction

Contributing

validate_datasets passing

The content of this repository can also be viewed here:

<https://bids-standard.github.io/bids-examples/>

bids-examples

This repository contains a set of **BIDS-compatible** datasets with **empty raw data files**. These datasets can be useful to:

1. write lightweight software tests
2. serve as an example on how a BIDS dataset can be structured

ALL RAW DATA FILES IN THIS REPOSITORY ARE EMPTY!

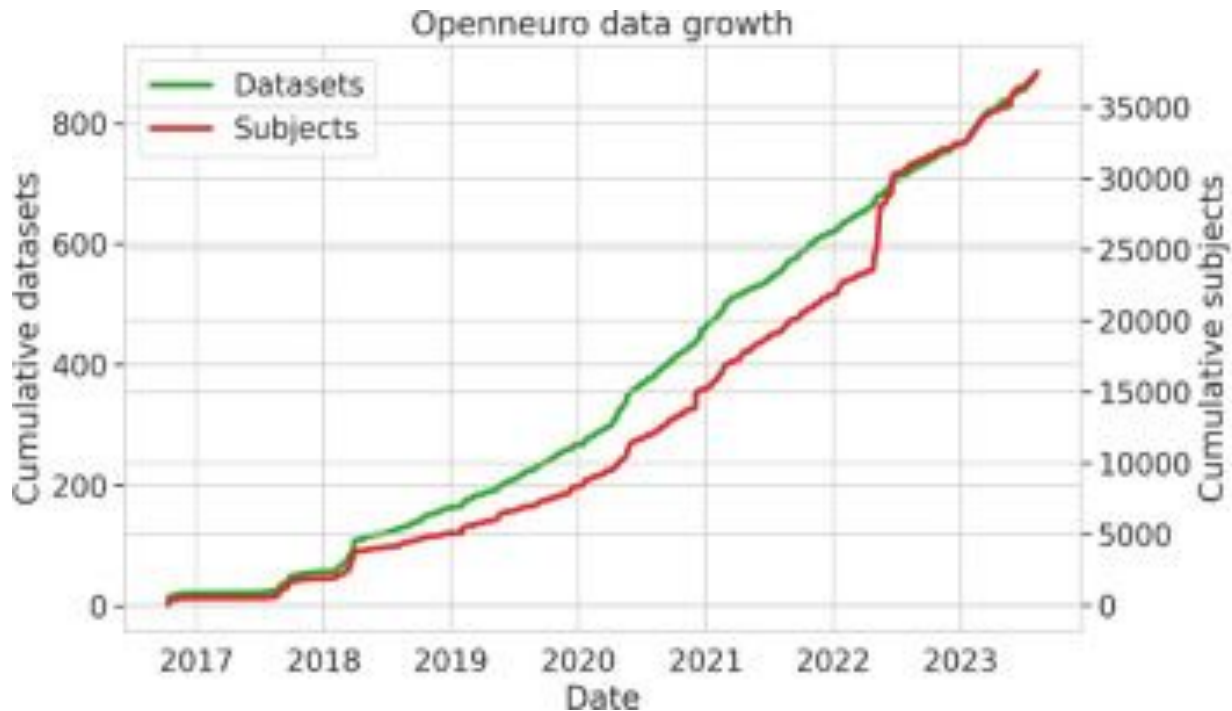
However for some of the data, the headers containing the metadata are still intact. (For example the NIFTI headers for `.nii` files, the BrainVision data headers for `.vhdr` files, or the OME-XML headers for `.ome.tif` files.)

Table of contents

- Validating BIDS examples
 - Validating individual examples
 - Validating all examples
 - Validator exceptions
- Contributing
- Dataset index
 - ASL
 - EEG
 - iEEG
 - MEG
 - Microscopy
 - Motion
 - MRI
 - NIRS
 - PET
 - qMRI

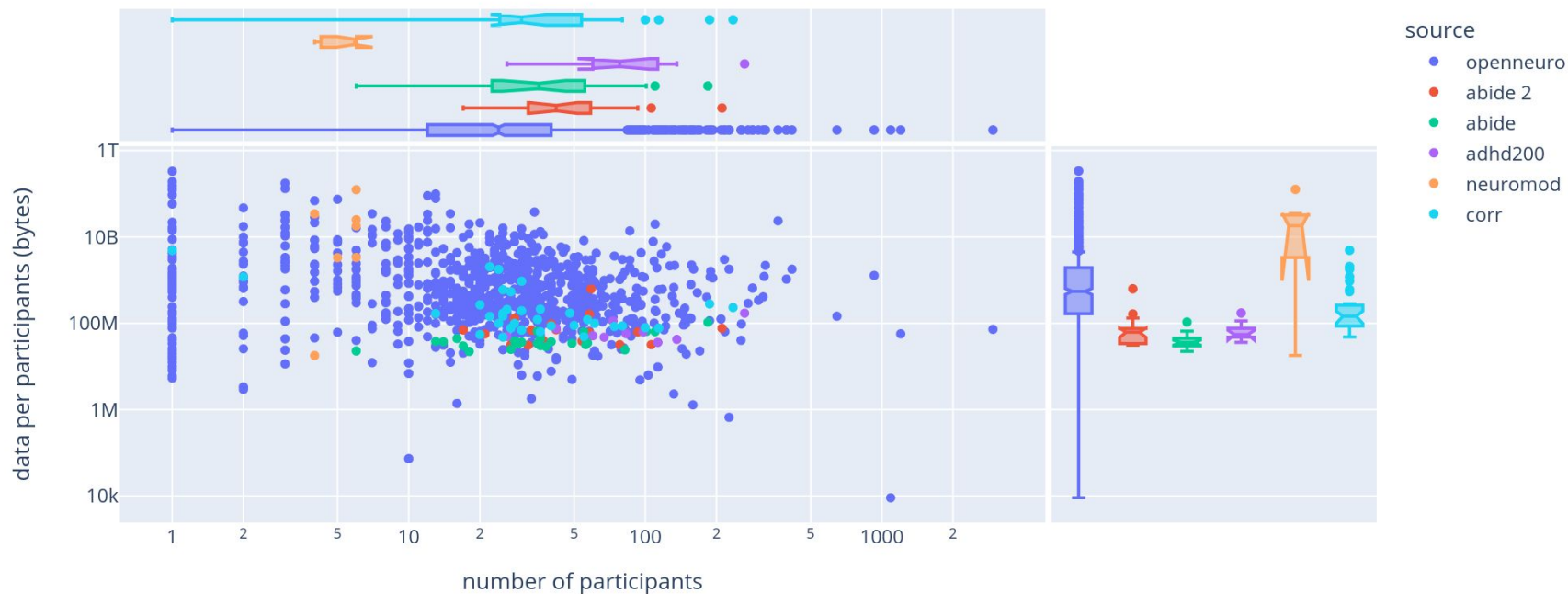
BIDS datasets

- [Openneuro](#)



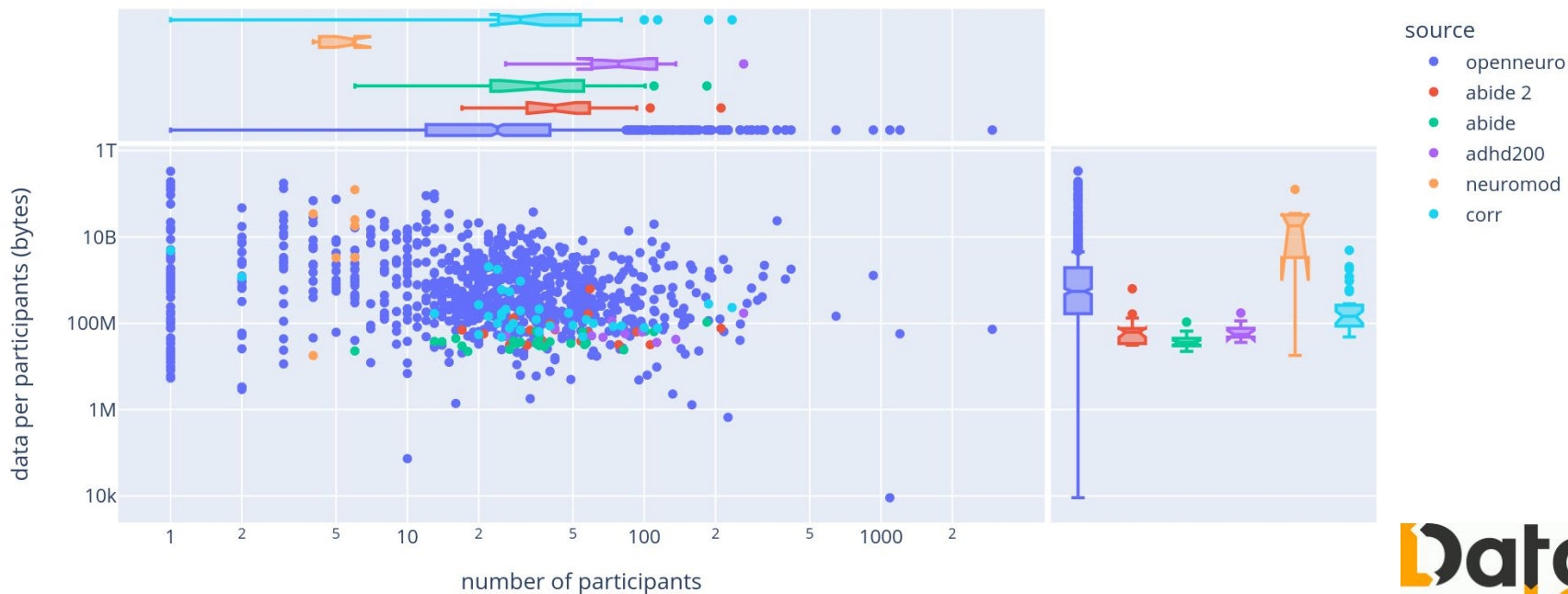
BIDS datasets

- Openneuro



BIDS datasets

- Openneuro



BIDS datasets

- [Neurobagel](#)

An ecosystem for distributed dataset harmonization and search.

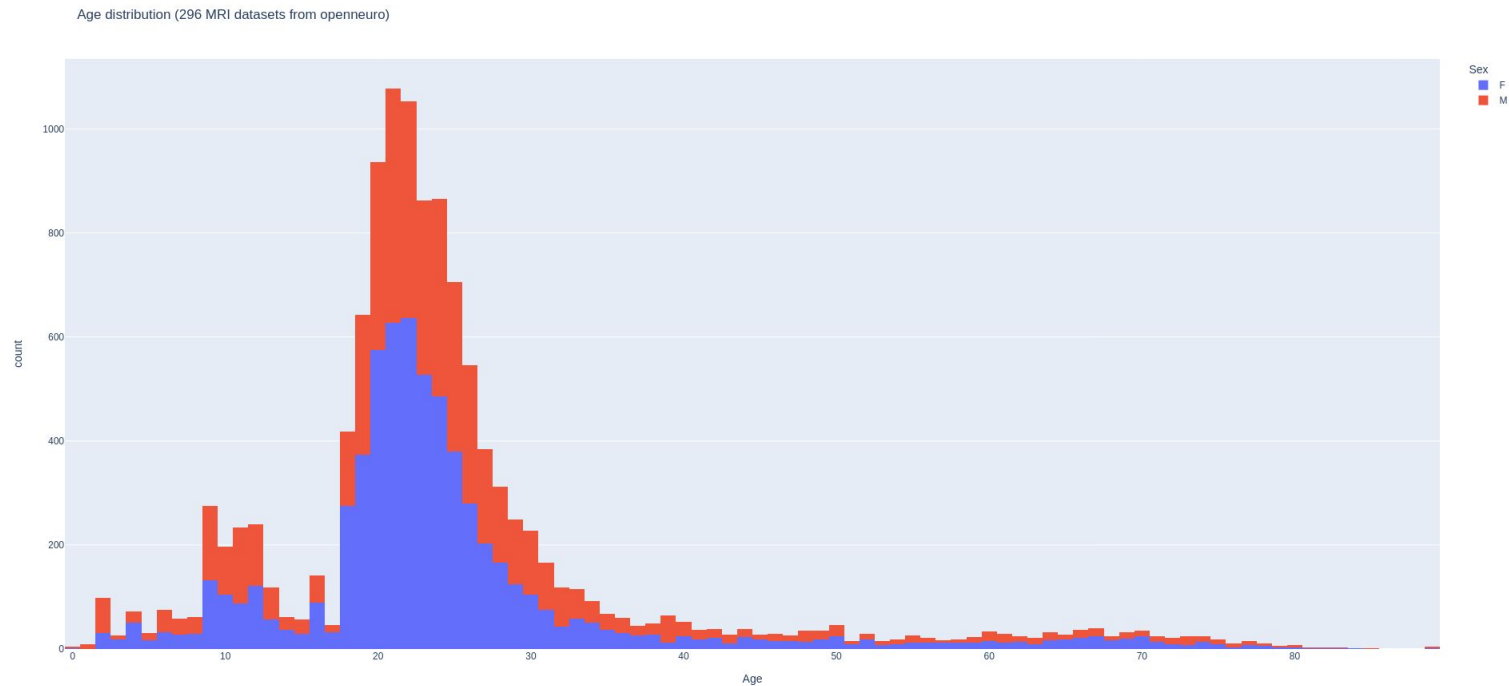
Neurobagel allows you to connect a local neuroimaging dataset with others in a decentralized framework using linked data principles.

Get started



BIDS datasets

- Neurobagel

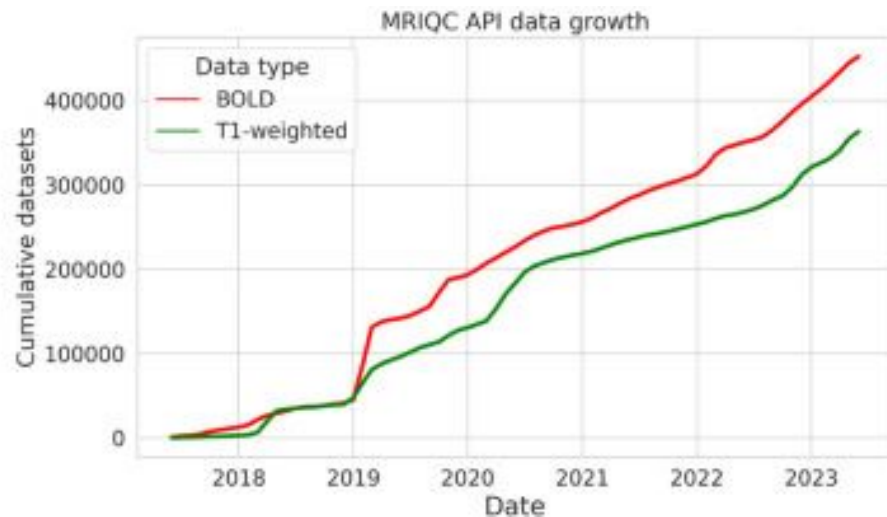
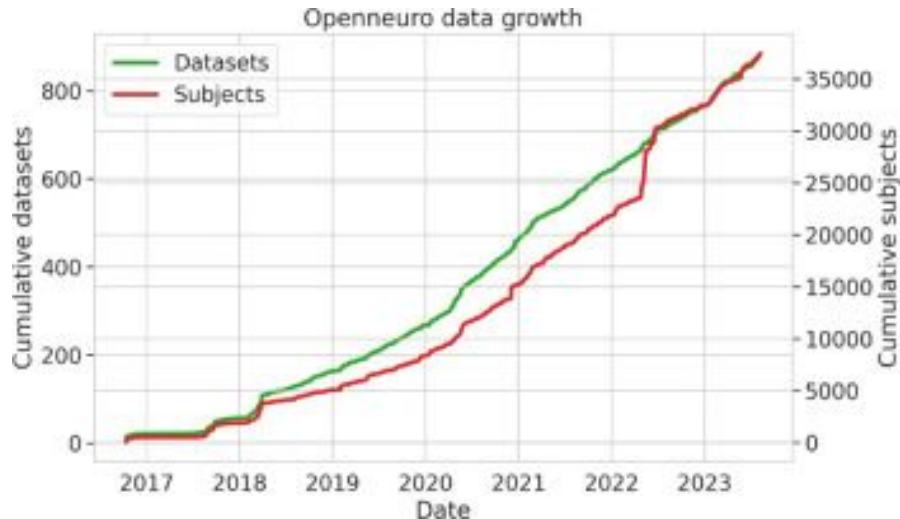


Using BIDS

- Standardization underlies automation: [BIDS apps](#)

Using BIDS

- Standardization underlies automation: [BIDS apps](#)




BIDS community

Questions:

- [Neurostars](#)
- BIDS [mailing list](#)
- Github [issues](#)
- [Mattermost](#)

Contact:

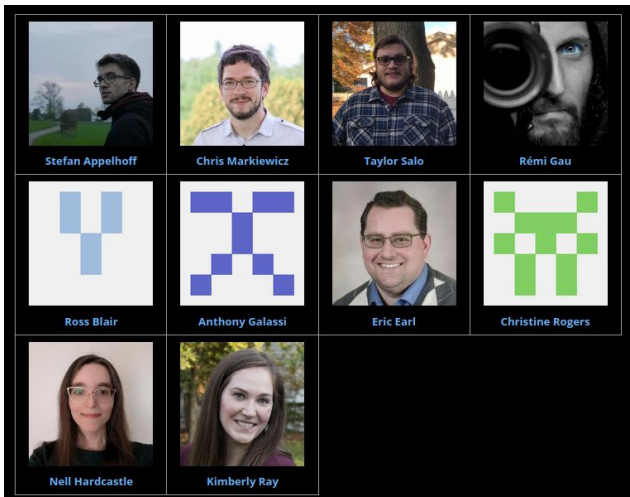
- X: <https://twitter.com/BIDSstandard>
-  <https://fosstodon.org/@bidsstandard>
- Bluesky: <https://bsky.app/profile/bidsstandard.bsky.social>

BIDS community

- >300 contributors!

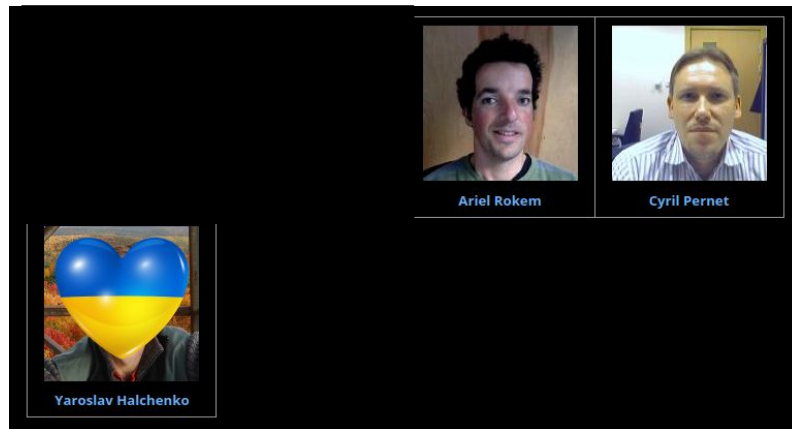
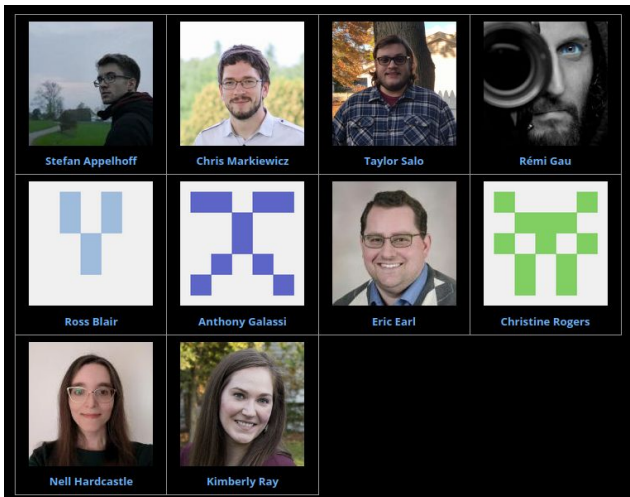
BIDS community

- [>300 contributors!](#)
- The BIDS steering group
- The BIDS maintainers



BIDS community

- >300 contributors!
- The BIDS steering group
- The BIDS maintainers



- Camille Maumet
- Dora Hermes

THANKS

<https://neurobagel.org>



Nikhil Bhagwat



Alyssa Dai



Arman Jahanpour



Jean-Baptiste Poline



Sebastian Urchs



<https://neurodatascience.github.io>

Why standardize data?

1. Das Fußmaß.

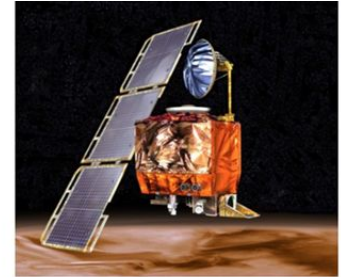
Der Wiener Fuß enthält 0,316102 Meter.

| Rahmen der Örter und ihrer Fußmaße. | Länge in Wien. Fuß | Länge in Metern |
|--|--------------------|-----------------|
| Baden, Fuß | 0,949 | 0,300 |
| Baiern, Fuß | 0,923 | 0,292 |
| Belgien, Aune (Meter) | 3,163 | 1,000 |
| Böhmen, Fuß | 0,938 | 0,296 |
| Dänemark, Fuß | 0,993 | 0,314 |
| England, Fuß (Foot) | 0,964 | 0,305 |
| Frankreich, Meter | 3,163 | 1,000 |
| alter Pariser Fuß | 1,028 | 0,325 |
| Griechenland, alte Piki (Endrezeh) | 2,050 | 0,648 |
| Hamburg, Fuß | 0,906 | 0,286 |
| Hannover, Fuß | 0,924 | 0,292 |
| Holland, Palm | 0,316 | 0,100 |
| Lemberg, Fuß | 0,939 | 0,297 |
| Neapel, Palmo | 0,834 | 0,264 |
| Nordamerikanische Freystaaten, 1 Foot | 0,964 | 0,305 |
| Polen, Fuß (Stopa) | 0,911 | 0,288 |
| Portugal, Fuß | 1,040 | 0,329 |
| neue Vara (Meter) | 3,163 | 1,000 |
| Preußen, Fuß | 0,993 | 0,314 |
| Rom, Fuß | 0,942 | 0,298 |
| Rußland, Saſchen | 6,750 | 2,134 |
| ruffiſcher Fuß | 0,964 | 0,305 |
| Sachſen, Fuß | 0,896 | 0,283 |
| Sardinien, Metro | 3,163 | 1,000 |

Source: [Wikipedia](#)

LISA GROSSMAN 11.10.10 7:00 AM

NOV. 10, 1999: METRIC MATH MISTAKE MUFFED MARS METEOROLOGY MISSION



The **\$125 million satellite** was supposed to be the first weather observer on another world. But as it approached the red planet to slip into a stable orbit Sept. 23, the [orbiter vanished](#). Scientists realized quickly it was gone for good. “It was pretty clear that morning, within half-an-hour, that the spacecraft had more or less **hit the top of the atmosphere and burned up**,” recalled NASA engineer Richard Cook, who was project manager for Mars exploration projects at the time.

A NASA review board found that the problem was in the software controlling the orbiter’s thrusters. **The software calculated the force the thrusters needed to exert in pounds of force. A separate piece of software took in the data assuming it was in the [metric unit: newtons](#).**

<https://www.wired.com/2010/11/1110mars-climate-observer-report/>

Source: [The Turing Way](#)

File naming

- Avoid white spaces
 - They make scripting harder.
- Use only letters, numbers, hyphens, and underscores.
 - Some OS can't handle special characters.
- Don't rely on letter case
 - For some OS "a" is the same as "A".
- Use separators in a meaningful way
 - ThisIsCamelCase
 - this_is_snake_case
 - This-is-kebab-case

Source: [Datalad RDM course](#)

File formats: metadata

sub-01_ses-post_task-livingnonlivingdecisionwithplainormirrorreversedtext_run-01_bold.nii.gz

```
{ "TaskName": "living-nonliving decision with plain or mirror-reversed text",  
  "RepetitionTime": 2.0}
```

sub-01_ses-post_task-decision_run-01_bold.nii.gz

```
{ "TaskName": "decision",  
  "RepetitionTime": 2.0,  
  "TaskDescription": "Subjects are presented with words in either plain text or mirror-reversed format,  
and are asked to judge whether the stimulus refers to a living or nonliving object. Items are presented in  
a mixed fashion and separated by whether each stimulus is a switch in presentation form from the previous  
trial.",  
  "CogAtlasID": "https://www.cognitiveatlas.org/id/trm\_5176cf9d3d512/",  
  "Instructions": "Subjects were told to ..."  
}
```

Minimize duplication

“I have 40 subjects, with 2 sessions and 4 runs per subject per session. I have 320 JSON files to edit. HELP!”

```
MyBidsProject
├── sub-control01
│   └── func
│       ├── sub-control01_task-nback_bold.json
│       └── sub-control01_task-nback_bold.nii
└── sub-control02
    └── func
        ├── sub-control02_task-nback_bold.json
        └── sub-control02_task-nback_bold.nii
```

```
MyBidsProject
├── sub-control01
│   └── func
│       └── sub-control01_task-nback_bold.nii
├── sub-control02
│   └── func
│       └── sub-control02_task-nback_bold.nii
└── task-nback_bold.json
```

[“Don’t repeat yourself” \(DRY\) principle](#)

In BIDS: [Inheritance principle](#)

Extending BIDS

- “*My data type is not yet part of BIDS.*”
 - Is it already in a [BIDS extension proposal](#)?
 - As google doc (e.g. [brain stimulation](#))
 - As pull request on the BIDS specification (e.g. [eye-tracking](#))

BIDS datasets

BIDS datasets

Number of datasets: 1113 with 48579 subjects

Total amount of data: 69 TB

Openneuro datasets

Number of datasets: 1018 with 43733 subjects

Total amount of data: 67 TB

With participants.tsv: 71 %

- 781 datasets with MRI data
- with fmriprep: 97 (4179 subjects)
- with freesurfer: 38 (3342 subjects)
- with mriqc: 333 (14829 subjects)

Using BIDS

- Basic tools to work with BIDS
 - Python: [pybids](#)
 - MATLAB / Octave: [bids-matlab](#)
 - R: [bidser](#)
- Automated tools: [BIDS apps](#)

Annotating your data

- minimally valid dataset + automation:
 - easy to generate statistical maps
- without good metadata:
 - impossible to interpret them

Enrich your dataset with more contextual information to help in the interpretation of the results.